Statistik für Informatiker, SS 2019

1 Grundlagen aus der Wahrscheinlichkeitstheorie

1.3 Bedingte Wahrscheinlichkeiten, Unabhängigkeit, gemeinsame Verteilung

Matthias Birkner

http://www.staff.uni-mainz.de/birkner/StatInfo19/

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

13.5.2019 / 20.5.2019

Ein Beispiel zur Motivation

Beispiel 1.37

Wir ziehen zwei Kugeln ohne Zurücklegen aus einer Urne mit $s > 0$ schwarzen und $w > 0$ weißen Kugeln. Wir stellen uns die Kugeln nummeriert vor, Nr. $1, \dots, w$ seien weiß, $w+1, \dots, w+s$ schwarz. Sei X_i die Nr. der Kugel im i-ten Zug ($= 1, 2$).

Also: $X = (X_1, X_2)$ ist uniform verteilt auf $S = \big\{(i,j) : 1 \le i, j \le w+s, i \neq j\big\}$.
Betrachte die Ereignisse

$$A = \big\{\text{erste Kugel ist weiß}\big\} = \{X_1 \le w\},$$

$$B = \big\{\text{zweite Kugel ist weiß}\big\} = \{X_2 \le w\}$$

$$\Big(= \big\{X \in \{(i,j) \in S : j \le w\}\big\}\Big)$$

Beispiel 1.37, Forts.

$$A = \{\text{erste Kugel ist weiß}\}, \quad B = \{\text{zweite Kugel ist weiß}\}$$

Ohne weitere Informationen ist

$$P(B) = P\big(X \in \{(i,j) \in S : j \le w\}\big)$$
$$= \frac{\#\{(i,j) \in S : j \le w\}}{\#S} = \frac{w(w+s-1)}{(w+s)(w+s-1)} = \frac{w}{w+s}.$$

(und übrigens $= P(A)$)

Nehmen wir an, wir haben den ersten Zug beobachtet und gesehen, dass A eingetreten ist. Mit dieser Information sollte die W'keit von B

$$\frac{w-1}{w+s-1} < \frac{w}{w+s}$$

sein

(denn es „wurde schon eine weiße Kugel verbraucht").

Beobachtung und Definition 1.38

Sei $(\Omega, \mathcal{F}, P)$ *W'raum,* $A \in \mathcal{F}$ *mit* $P(A) > 0$*. Für* $B \in \mathcal{F}$

$$P(B \mid A) := \frac{P(B \cap A)}{P(A)}$$

heißt bedingte Wahrscheinlichkeit von B, gegeben A.

$P(\cdot \mid A)$ *ist ein Wahrscheinlichkeitsmaß auf* Ω*, d.h. die Eigenschaften aus Forderung (1.2) Normierung und* σ*-Additivität sind erfüllt. (Prüfung per Inspektion)*

Wir lassen $P(B \mid A)$ undefiniert, wenn $P(A) = 0$.

In Beispiel 1.37 ist

$$A = \{\text{erste Kugel ist weiß}\}, \quad B = \{\text{zweite Kugel ist weiß}\}$$

$$P(A) = \frac{w}{w+s},$$
$$P(A \cap B) = P(X_1 \le w, X_2 \le w) = \frac{w(w-1)}{(w+s)(w+s-1)},$$

also ergibt sich tatsächlich $P(B \mid A) = \dfrac{w-1}{w+s-1}$.

Bemerkung 1.39 („Natürlichkeit von Definition 1.38“)

Nehmen wir an, wir möchten angesichts der Information „A ist eingetreten“ das W'maß P revidieren zu einem W'maß $\widetilde{P}$ mit

1. $\widetilde{P}(A) = 1$ (d.h. A ist sicher unter $\widetilde{P}$) und
2. $\widetilde{P}(B) = c_A P(B)$ für $B \subset A$ mit einem $c_A > 0$ (d.h. Teilereignisse von A erhalten bis auf Normierung ihr altes Gewicht).

Dann gilt

$$\widetilde{P}(C) = \frac{P(A \cap C)}{P(A)} \ \big(= P(C \mid A) \big) \quad \text{für alle } C \in \mathcal{F}.$$

Beweis: Für $C \in \mathcal{F}$ ist $\widetilde{P}(C) = \widetilde{P}(A \cap C) + \underbrace{\widetilde{P}(C \setminus A)}_{\leq \widetilde{P}(A^c) = 0} \stackrel{2.}{=} c_A P(A \cap C)$,

mit Wahl $C = A$ und 1. folgt $1 = \widetilde{P}(A) = c_A P(A)$, also $c_A = 1/P(A)$.

Beachte: (Stochastische) Abhängigkeit und Kausalität sind nicht dasselbe

Bemerkung 1.40

$P(B \mid A) \neq P(B)$ kann nicht notwendigerweise als „Kausalität“ (im Sinne von „A beeinflusst, ob B eintritt“) interpretiert werden:

So ist in Beispiel 1.37 ist auch

$$P(A \mid B) = \frac{P(B \cap A)}{P(B)} = \frac{w-1}{w+s-1} \neq P(A),$$

aber es passt nicht zu unserer Vorstellung, dass der 2. Zug den 1. Zug beeinflusst.

Satz 1.41

Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, I eine (höchstens abzählbare) Indexmenge, $B_i \in \mathcal{F}$ paarweise disjunkt mit $P\left(\bigcup_{i\in I} B_i\right) = 1$ und $P(B_i) > 0$ für $i \in I$.

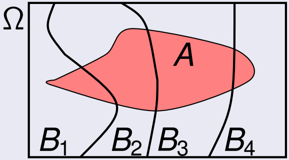
1. (Formel von der totalen Wahrscheinlichkeit)

Für $A \in \mathcal{F}$ gilt

$$P(A) = \sum_{i\in I} P(B_i)P(A \mid B_i).$$

Man kann dies anhand eines Diagramms veranschaulichen:

$P(A \mid B_i)$ ist der Anteil von $A \cap B_i$ an der „Wahrscheinlichkeitsmasse" $P(B_i)$ von B_i



Beweis:

$$\sum_{i\in I} P(B_i)P(A \mid B_i) = \sum_{i\in I} P(A \cap B_i) = P\left(\bigcup_{i\in I} (A \cap B_i)\right) = P(A)$$

Satz 1.41 (Fortsetzung)

$B_i \in \mathcal{F}$ paarw. disjunkt mit $P\left(\bigcup_{i \in I} B_i\right) = 1$ und $P(B_i) > 0$ für $i \in I$

2. (Formel von Bayes)

Für $A \in \mathcal{F}$ mit $P(A) > 0$ und jedes $k \in I$ gilt

$$P(B_k \mid A) = \frac{P(B_k)P(A \mid B_k)}{\sum_{i \in I} P(B_i)P(A \mid B_i)}$$

Insbesondere gilt für Ereignisse A, B mit $P(A) > 0$

$$P(B \mid A) = \frac{P(B)P(A \mid B)}{P(B)P(A \mid B) + P(B^c)P(A \mid B^c)}$$

(verwende Zerlegung $B \cup B^c = \Omega$)

Beweis: Der Nenner ist $= P(A)$ nach 1.,
der Zähler ist $= P(A \cap B_k)$ nach Def.

Beispiel 1.42 (Verrauschter Übertragungskanal)

Ein (einzelnes) Bit X (aus $\{0,1\}$, es gelte $P(X=1)=a\in(0,1)$) wird über einen fehleranfälligen Kanal gesendet, der jede 1 mit W'keit f_1 und jede 0 mit W'keit f_0 flippt, sei Y das empfangene Bit.

Dann ist

$$P(Y=0\,|\,X=1)=f_1,\quad P(Y=1\,|\,X=1)=1-f_1,$$
$$P(Y=1\,|\,X=0)=f_0,\quad P(Y=0\,|\,X=0)=1-f_0,$$

also

$$\begin{aligned}&P(Y=1)\\&=P(X=1)P(Y=1\,|\,X=1)+P(X=0)P(Y=1\,|\,X=0)\\&=a(1-f_1)+(1-a)f_0,\end{aligned}$$

Beispiel 1.42 (Fortsetzung)

Ein (einzelnes) Bit X (aus $\{0, 1\}$, es gelte $P(X = 1) = a \in (0, 1)$) wird über einen fehleranfälligen Kanal gesendet, der jede 1 mit W'keit f_1 und jede 0 mit W'keit f_0 flippt, sei Y das empfangene Bit.

Analog ist

$$\begin{aligned} &P(Y = 0) \\ &= P(X = 0)P(Y = 0 \mid X = 0) + P(X = 1)P(Y = 0 \mid X = 1) \\ &= (1 - a)(1 - f_0) + af_1 \end{aligned}$$

Beispiel 1.42, Forts.

$$P(X = 1 \mid Y = 1)$$
$$= \frac{P(X = 1)P(Y = 1 \mid X = 1)}{P(X = 1)P(Y = 1 \mid X = 1) + P(X = 0)P(Y = 1 \mid X = 0)}$$
$$= \frac{a(1 - f_1)}{a(1 - f_1) + (1 - a)f_0}$$

$$P(X = 0 \mid Y = 0)$$
$$= \frac{P(X = 0)P(Y = 0 \mid X = 0)}{P(X = 0)P(Y = 0 \mid X = 0) + P(X = 1)P(Y = 0 \mid X = 1)}$$
$$= \frac{(1 - a)(1 - f_0)}{(1 - a)(1 - f_0) + af_1}$$

Beispiel 1.42 (Fortsetzung)

Ein (einzelnes) Bit X (aus $\{0, 1\}$, es gelte $P(X = 1) = a \in (0, 1)$) wird über einen fehleranfälligen Kanal gesendet, der jede 1 mit W'keit f_1 und jede 0 mit W'keit f_0 flippt, sei Y das empfangene Bit.

Z.B. für die konkreten Werte $a = 0.3$, $f_1 = 0.05$, $f_0 = 0.1$ ergibt sich $P(Y = 1) = 0.355$

$$P(X = 1 \mid Y = 1) \approxeq 0.803, \quad P(X = 0 \mid Y = 0) \approxeq 0.977.$$

Beispiel 1.43 (Medizinische Reihenuntersuchung)

Eine Krankheit

- komme bei 2% der Bevölkerung vor („Prävalenz 2%"),
- ein Test schlage bei 95% der Kranken an („Sensitivität 95%"),
- aber auch bei 10% der Gesunden („Spezifität 90%").

Eine zufällig gewählte Person wird mit positivem Resultat getestet. Wie wahrscheinlich ist es, dass sie tatsächlich krank ist?

Sei

$$A := \{\text{Test fällt positiv aus}\}, \quad B := \{\text{getestete Person ist krank}\},$$

also $P(B) = 0{,}02$, $P(A \mid B) = 0{,}95$, $P(A \mid B^c) = 0{,}1$, somit

$$\begin{aligned} P(B \mid A) &= \frac{P(B)P(A \mid B)}{P(B)P(A \mid B) + P(B^c)P(A \mid B^c)} \\ &= \frac{0{,}02 \cdot 0{,}95}{0{,}02 \cdot 0{,}95 + 0{,}98 \cdot 0{,}1} \approxeq 0{,}162 \end{aligned}$$

(wir sehen: geringe „positive Korrektheit“), andererseits

$$\begin{aligned} P(B \mid A^c) &= \frac{P(B)P(A^c \mid B)}{P(B)P(A^c \mid B) + P(B^c)P(A^c \mid B^c)} \\ &= \frac{0{,}02 \cdot 0{,}05}{0{,}02 \cdot 0{,}05 + 0{,}98 \cdot 0{,}9} \approxeq 0{,}0011 \end{aligned}$$

(recht hohe „negative Korrektheit“).

Demnach: Ein negatives Testergebnis schließt die Krankheit mit hoher W'keit aus, ein positives Testergebnis sollte eher als „der Fall sollte weiter beobachtet / untersucht werden" interpretiert werden als „die Testperson ist krank".

S.a. Gerd Gigerenzer, *Das Einmaleins der Skepsis*, Berlin Verlag, 2002, der auch einlädt, den Sachverhalt anschaulich anhand einer „Vierfelder-Tafel" bezogen auf eine Gesamtpopulation der Größe 1000 zu betrachten:

	krank	gesund	Σ
pos. getestet	19	98	117
neg. getestet	1	882	883
Σ	20	980	1000

Beobachtung 1.44 (Multiplikationsformel)

$A_1, A_2, \dots, A_n \in \mathcal{F}$ Ereignisse (auf einem W'raum $(\Omega, \mathcal{F}, P)$ erklärt) mit

$$P(A_1 \cap \dots \cap A_{n-1}) > 0,$$

so ist

$$\begin{aligned} &P(A_1 \cap A_2 \cap \dots \cap A_n) \\ &= P(A_1) P(A_2 \mid A_1) P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap \dots \cap A_{n-1}). \end{aligned}$$

(Beweis per Inspektion, das Produkt rechts teleskopiert)

Betrachten wir z.B. in der Situation von Bsp. 1.37 (Ziehen ohne Zurücklegen aus Urne mit s schwarzen und w weißen Kugeln) drei Züge, sei
$A_1 = \{\text{erste Kugel ist weiß}\}$, $A_2 = \{\text{zweite Kugel ist weiß}\}$,
$A_3 = \{\text{dritte Kugel ist schwarz}\}$, so ist

$$P(A_1) = \frac{w}{s+w}, \ P(A_2 \mid A_1) = \frac{w-1}{s+w-1},$$

$$P(A_3 \mid A_1 \cap A_2) = \frac{s}{s+w-2},$$

also

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= \frac{w}{s+w} \cdot \frac{w-1}{s+w-1} \cdot \frac{s}{s+w-2} \\ &= \frac{w(w-1)s}{(s+w)(s+w-1)(s+w-2)}. \end{aligned}$$

Gemeinsame Verteilung und Randverteilungen

Definition 1.45 (Gemeinsame Verteilung und Marginalverteilungen)

Seien $X_1, X_2, \dots, X_d$ ZVn (die auf einem gemeinsamen W'raum $(\Omega, \mathcal{F}, P)$ definiert sind), X_i habe Werte in S_i,

$$X := \big(X_1, X_2, \dots, X_d\big)$$

die Produkt-Zufallsvariable (mit Werten in $S_1 \times \cdots \times S_d$), so heißt $\mathscr{L}(X)$ die *gemeinsame Verteilung* der $X_1, X_2, \dots, X_d$ ($\mathscr{L}(X)$ ist ein Wahrscheinlichkeitsmaß auf dem Wertebereich von X, d.h. auf dem Produktraum $S_1 \times \cdots \times S_d$, vgl. Definition 1.10).
$\mathscr{L}(X_i)$ heißt die i-te *Randverteilung* (oder *Marginalverteilung*) von X.

Beispiel

Fassen wir die Häufigkeitstabelle (normiert auf Gesamtsumme = 1) aus Bsp. 1.43 als Verteilungstafel der gemeinsamen Verteilung von

X_1 („Testergebnis" mit Werten in $S_1 = \{\text{positiv}, \text{negativ}\}$) und

X_2 („Gesundheitszustand" mit Werten in $S_2 = \{\text{krank}, \text{gesund}\}$) auf:

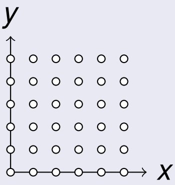
X_1 \ X_2	krank	gesund	
positiv	0.019	0.098	0.117
negativ	0.001	0.882	0.883
	0.02	0.98	1

Beispiel 1.46

1. Sei $X = (X_1, Y_1)$ uniform verteilt auf dem (diskreten) Quadrat $Q_L := \{0, 1, \dots, L-1\}^2 = \{(x, y) \in \mathbb{N}_0^2 : \max\{x, y\} < L\}$ (mit $L \in \mathbb{N}$), d.h.

$$P(X_1 = x, Y_1 = y) = \frac{1}{\#Q_L} = \frac{1}{L^2} \quad \text{für } x, y \in \{0, 1, \dots, L-1\}.$$

Somit ist
$P(X_1 = x) = \sum_{y=0}^{L-1} P(X_1 = x, Y_1 = y) = \frac{1}{L}$
für $0 \le x < L$ und ebenso
$P(Y_1 = y) = \frac{1}{L}$ für $0 \le y < L$.



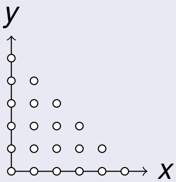
Wir sehen hier:
$P(X_1 = x, Y_1 = y) = P(X_1 = x) \cdot P(Y_1 = y)$, d.h. die gemeinsamen Gewichte erhält man als *Produkt* der Randverteilungs-Gewichte. Man sagt dazu auch: X_1 und Y_1 sind *unabhängig* (siehe Def. 1.51 unten für den allgem. Fall).

Beispiel 1.46 (Fortsetzung)

2. Sei $X = (X_2, Y_2)$ uniform verteilt auf dem (diskreten) Dreieck $D_L := \{(x, y) \in \mathbb{N}_0^2 : x + y < L\}$, d.h.

$$P(X_2 = x, Y_2 = y) = \frac{1}{\#D_L} = \frac{2}{L(L+1)} \quad \text{für } (x, y) \in D_L$$

Somit ist $P(X_2 = x) = \sum_y P(X_2 = x, Y_2 = y) = \frac{L-x}{\#D_L} = \frac{2(L-x)}{L(L+1)}$ für $0 \leq x < L$ und analog $P(Y_2 = y) = \frac{2(L-y)}{L(L+1)}$ für $0 \leq y < L$.

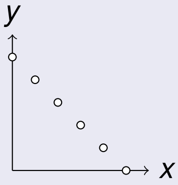


Wir sehen in diesem Fall:
$P(X_2 = x, Y_2 = y) \neq P(X_2 = x) \cdot P(Y_2 = y)$, die Koordinaten X_2 und Y_2 sind nicht unabhängig.

Beispiel 1.46 (Fortsetzung)

3. Sei $X = (X_3, Y_3)$ uniform verteilt auf der (diskreten, verschobenen) Nebendiagonale $N_L := \{(x, y) \in \mathbb{N}_0^2 : x + y = L\}$, d.h.

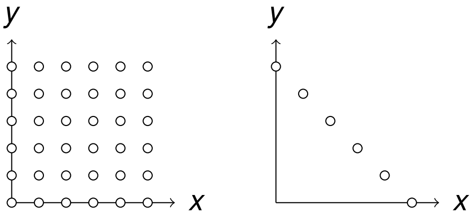
$$P(X_3 = x, Y_3 = y) = \frac{1}{\# N_L} = \frac{1}{L} \quad \text{für } (x, y) \in N_L$$



Somit ist
$P(X_3 = x) = \sum_y P(X_3 = x, Y_3 = y) = \frac{1}{L}$ für $0 \le x < L$ und analog
$P(Y_3 = y) = \frac{1}{L}$ für $0 \le y < L$.

Wir sehen in diesem Fall:
$P(X_3 = x, Y_3 = y) \neq P(X_3 = x) \cdot P(Y_3 = y)$, die Koordinaten X_3 und Y_3 sind nicht unabhängig.
(Es gilt hier sogar $Y_3 = L - X_3$, d.h. es gibt einen deterministischen Zusammenhang.)

Zudem sehen wir an Beispiel 1.46, 1. und 3. zusammen:
Die Randverteilungen legen die gemeinsame Verteilung nicht fest.



So gilt $\mathscr{L}(X_1) = \mathscr{L}(X_3)$ und $\mathscr{L}(Y_1) = \mathscr{L}(Y_3)$ (man schreibt dies auch als $X_1 =^d X_3$, $Y_1 =^d Y_3$, „Gleichheit in Verteilung"), aber

$$\mathscr{L}\big((X_1, Y_1)\big) \neq \mathscr{L}\big((X_3, Y_3)\big).$$

Bedingte Verteilung und mehrstufige Experimente

Wir haben ZVn $X_1, X_2, \ldots, X_n$ im Sinn und kennen

1. die Verteilung von X_1,
2. für $2 \le k \le n$ die bedingte Verteilung von X_k, wenn $X_1, X_2, \ldots, X_{k-1}$ schon beobachtet wurden.
   (d.h. für beliebige beobachtete Werte $x_1, x_2, \ldots, x_{k-1}$ kennen wir $P(X_k \in \cdot \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1})$)

Dann kann man die gemeinsame Verteilung (zumindest im diskreten Fall, d.h. die gemeinsamen Gewichte) des Vektors $(X_1, X_2, \ldots, X_n)$ mittels der Multiplikationsformel (Beob. 1.44) bestimmen („Pfadregel"):

$$
\begin{aligned}
&P(X_1 = x_1, X_2 = X_2, \ldots, X_n = x_n) \\
&= P(X_1 = x_1) \cdot P(X_2 = x_2 \mid X_1 = x_1) \cdot P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2) \\
&\qquad \cdot \cdots \cdot P(X_n = x_n \mid X_1 = x_1, X_2 = x_2, \ldots, X_{n-1} = x_{n-1})
\end{aligned}
$$

Man stellt Rechnungen, die die verschiedenen Fälle („längs der „Pfade") in dieser Weise aufzählen, oft mittels eines Baumdiagramms dar, wie in dem folgenden Beispiel.

Beispiel 1.47

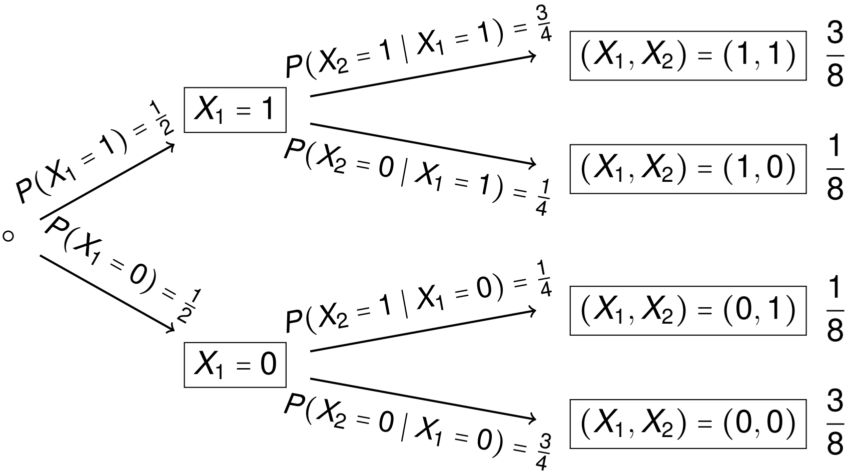
Wir haben eine faire Münze M_1 und zwei gezinkte Münzen M_2, M_3 (deren Seiten jeweils mit 0 und 1 beschriftet seien), wobei

$$P(M_2 = 1) = \frac{3}{4}, P(M_3 = 1) = \frac{1}{4}.$$

Wir werfen erst M_1, wenn M_1 den Wert 1 zeigt, so werfen wir dann M_2, sonst M_3. Sei

$$X_i = \text{Resultat des } i\text{-ten Wurfs}, \; i = 1, 2.$$

Beispiel 1.47, Forts.



Beispiel 1.47, Forts.

Die gemeinsame Verteilung von (X_1, X_2) ist damit

X_1 \ X_2	1	0	
1	$\frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$	$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$	$\frac{1}{2}$
0	$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$	$\frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

Wir sehen hier wieder: Die Randverteilungen legen (i.A.) nicht die gemeinsame Verteilung fest. Es ist
$P(X_1 = 1) = P(X_2 = 1) = \frac{1}{2}$
(und dieselben Randverteilungen ergäben sich, wenn man zwei Mal M_1 wirft, aber die gemeinsame Verteilung wäre eine andere).

Beispiel 1.48

Nehmen wir an, in der Situation von Bsp. 1.42 („verrauschter Übertragungskanal") wird das zufällige Bit X sicherheitshalber zweimal gesendet (wobei jedesmal unabhängig mit den genannten W'keiten ein Übertragungsfehler auftritt), seien Y_1 und Y_2 die beiden empfangenen Bits, $Z_1 = 1_{\{Y_1=Y_2\}}$, $Z_2 = 1_{\{Y_1=Y_2=X\}}$ (beachte, dass der Empfänger Z_1 beobachten kann, nicht aber Z_2).

Dann ist wegen $\{Z_2 = 1\} \subset \{Z_1 = 1\}$

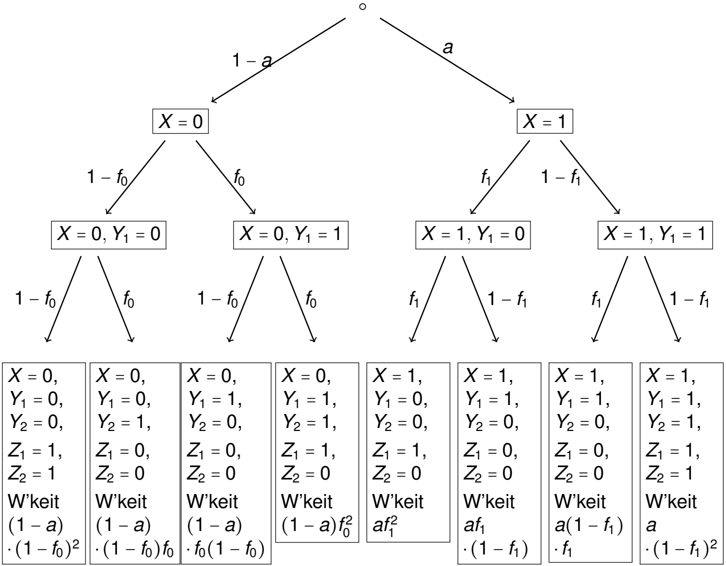
$$P(Z_2 = 1 \mid Z_1 = 1) = \frac{P(Z_2 = 1)}{P(Z_1 = 1)}$$

(dies ist die Wahrscheinlichkeit, mit der ein Empfänger, der den gesendeten Bits „vertraut", sofern er zweimal dasselbe empfangen hat, „richtig liegt").

Beispiel 1.48, Forts.

Wir fassen die möglichen Ausgänge von (X, Y_1, Y_2) (und die sich daraus ergebenden Werte von Z_1, Z_2) in einem Baumdiagramm zusammen:

Beispiel 1.48, Forts.



Beispiel 1.48, Forts.

Wir sehen:

$$P(Z_1 = 1) = (1-a)(1-f_0)^2 + (1-a)f_0^2 + af_1^2 + a(1-f_1)^2$$

$$P(Z_2 = 1) = (1-a)(1-f_0)^2 + a(1-f_1)^2$$

$$P(Z_2 = 1 \mid Z_1 = 1) = \frac{(1-a)(1-f_0)^2 + a(1-f_1)^2}{(1-a)(1-f_0)^2 + (1-a)f_0^2 + af_1^2 + a(1-f_1)^2}$$

Für die konkreten Zahlenwerte aus Beispiel 1.42 ($a = 0.3$, $f_1 = 0.05$, $f_0 = 0.1$) ergibt sich

$$P(Z_2 = 1 \mid Z_1 = 1) \approxeq 0.991.$$

Unabhängigkeit

Definition 1.49

Ereignisse A und B (auf demselben W'raum) heißen (stochastisch) *unabhängig*, falls gilt

$$P(A \cap B) = P(A)P(B)$$

(Sofern $P(A) > 0$, so gilt dann auch $P(B \mid A) = P(B)$, d.h. die bedingte und die unbedingte W'keit von B stimmen überein, analog mit vertauschten Rollen.)

Allgemein: Ereignisse $A_1, A_2, \dots, A_n$ heißen (stochastisch) *unabhängig*, falls gilt

$$P\Big(\bigcap_{i \in J} A_i\Big) = \prod_{i \in J} P(A_i) \quad \text{für alle } \varnothing \neq J \subset \{1, 2, \dots, n\} \qquad (1)$$

(für unendlich viele Ereignisse $A_1, A_2, \dots$: (1) für jedes endliche $J \subset \mathbb{N}$).

Beispiel

Wir ziehen eine Karte (nennen wir sie K) aus einem gut gemischtem Skatblatt, sei $A = \{K \text{ ist ein As}\}$, $B = \{K \text{ ist kreuz}\}$.

Es ist $P(A) = \frac{4}{32} = \frac{1}{8}$, $P(B) = \frac{8}{32} = \frac{1}{4}$, $P(A \cap B) = \frac{1}{32} = \frac{1}{8} \cdot \frac{1}{4} = P(A) \cdot P(B)$, d.h. A und B sind unabhängig.

Aber auch: Ziehe eine zweite Karte K_2 (ohne K zurückzustecken), sei $C = \{K_2 \text{ ist kreuz}\}$. Es ist $P(C) = \frac{8 \cdot 7 + 24 \cdot 8}{32 \cdot 31} = \frac{1}{4}$, $P(A \cap C) = \frac{3 \cdot 8 + 1 \cdot 7}{32 \cdot 31} = \frac{1}{32} = P(A) \cdot P(C)$,

d.h. auch A und C sind unabhängig (was vielleicht zumindest auf den ersten Blick nicht zur Intuition passt).

Die Ereignisse B und C sind *nicht* unabhängig.

Bemerkung 1.50

1. Sind Ereignisse $A_1, A_2, \ldots, A_n$ unabhängig, so gilt dies offenbar auch für jede Teilfamilie $A_{i_1}, \ldots, A_{i_k}$ (mit $1 \leq i_1 < \cdots < i_k \leq n$).

2. Es genügt i.A. nicht, in (1) nur Paare zu prüfen.

Beispiel: Wir werfen dreimal eine faire Münze, mit Ergebnissen W_1, W_2, W_3, sei $A_1 = \{W_1 = W_2\}$, $A_2 = \{W_1 = W_3\}$, $A_3 = \{W_2 = W_3\}$.
Es ist $P(A_i) = \frac{1}{2}$, $P(A_i \cap A_j) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} = P(A_i)P(A_j)$ für $i \neq j$, $i, j \in \{1, 2, 3\}$, aber
$P(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq P(A_1)P(A_2)P(A_3)$.

Man sagt dazu, dass die Ereignisse A_i hier *paarweise unabhängig* sind.

Diskussion.

1. Stochastische Unabhängigkeit ist eine (gemeinsame) Eigenschaft von Ereignissen und deren Wahrscheinlichkeiten; (Un-)abhängigkeit ist nicht automatisch mit (Nicht-)Existenz eines kausalen Zusammenhangs gleichzusetzen.

Diskussion.

Beispiel:

Wir befragen eine zufällig an einem Samstagnachmittag auf dem Mainzer Gutenbergplatz ausgewählte Testperson. Die Ereignisse „hat Schuhgröße ≥ 41" und „hat Führerschein" sind nicht unabhängig

(gegeben $\{$hat Schuhgröße $\geq 41\}$ handelt es sich vermutlich eher um einen Erwachsenen, daher ist die Chance, dass die Person auch einen Führerschein hat größer als der Anteil der Führerscheinbesitzer in der Gesamtbevölkerung, die auch viele Kinder umfasst).

Trotzdem wäre die Behauptung, dass große Füße Führerscheine hervorbringen, natürlich unsinnig.

Diskussion.

2. Nichtsdestoweniger *modelliert* man die erneute Wiederholung eines gewissen zufälligen Experiments unter gleichen Bedingungen (oder auch die Befragung verschiedener Versuchspersonen aus einer großen Grundgesamtheit) zumeist mittels (angenommener) stochastischer Unabhängigkeit.

Die Annahme unabhängiger Kopien eines gewissen Zufallsexperiments bildet häufig einen zentralen Ansatzpunkt statistischer Analysen.

Unabhängige Zufallsvariablen

Definition 1.51

Zufallsvariablen $X_1, X_2, \ldots, X_d$ (die auf einem gemeinsamen W'raum definiert sind) heißen (stochastisch) *unabhängig*, wenn für alle Ereignisse $\{X_i \in B_i\}$ gilt

$$P\big(X_1 \in B_1, X_2 \in B_2, \ldots, X_d \in B_d\big) = \prod_{i=1}^{d} P(X_i \in B_i) \tag{2}$$

Beobachtung 1.52

$X_1, X_2, \ldots, X_n$ ZVn, X_i habe Werte in S_i, S_i abzählbar für $i = 1, 2, \ldots, n$. Dann sind $X_1, X_2, \ldots, X_n$ unabhängig g.d.w. gilt $\forall\, x_1 \in S_1, x_2 \in S_2, \ldots, x_n \in S_n$:

$$P\big(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\big) = \prod_{i=1}^{n} P(X_i = x_i)$$

(d.h. die gemeinsamen Gewichte haben Produktgestalt).

Beweis: „$\Rightarrow$“: Klar (wähle $B_i = \{x_i\}$ in (2)).

„$\Leftarrow$“: Seien $B_1 \subset S_1, \ldots B_n \subset S_n$, es ist

$$P\big(\{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}\big) = P\Big(\bigcup_{(x_1,\ldots,x_n)\in B_1\times\cdots\times B_n} \{X_1 = x_1, \ldots, X_n = x_n\}\Big)$$

$$= \sum_{x_1\in B_1,\ldots,x_n\in B_n} \underbrace{P\big(X_1 = x_1, \ldots, X_n = x_n\big)}_{=P(X_1=x_1)\cdots P(X_n=x_n)} = \sum_{x_1\in B_1} P(X_1 = x_1)\cdots \sum_{x_n\in B_n} P(X_n = x_n)$$

$$= P(X_1 \in B_1)\cdots P(X_n \in B_n).$$

Beobachtung und Definition 1.53

Für unabhängige Zufallsvariablen $X_1, \dots, X_n$ ist also die gemeinsame Verteilung, d.h. die Verteilung von $X := \left(X_1, \dots, X_n\right)$ durch die Randverteilungen festgelegt (siehe Def. 1.45).

Man nennt dann $\mu := \mathscr{L}(X)$ das Produkt der $\mu_1 := \mathscr{L}(X_1), \dots, \mu_n := \mathscr{L}(X_n)$ und schreibt

$$\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$$

Bemerkung 1.54

Sind ZVn $X_1, \dots, X_n$ unabhängig, so auch

1. jede Teilfamilie $X_{i_1}, \dots, X_{i_k}$ (mit $1 \le i_1 < \dots < i_k \le n$) (wähle $B_i = S_i$ in (2) für $i \notin \{i_1, \dots, i_k\}$) ;

2. $f_1(X_1), f_2(X_2), \dots, f_n(X_n)$ für Funktionen $f_i : S_i \to S_i'$ (beachte $\{f_i(X_i) \in B_i'\} = \{X_i \in f_i^{-1}(B_i')\}$ in (2))

Bemerkung 1.54 (Fortsetzung)

3. In Def. 1.51 genügt es i.A. nicht, jeweils nur Paare auf Unabhängigkeit zu prüfen:

Beispiel (wir sprechen Bsp. 1.50, 2. mit ZVn aus):
Seien X_1, X_2, X_3 unabhängige faire Münzwürfe $P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$,
$Y_1 = 1_{\{X_1=X_2\}}$, $Y_2 = 1_{\{X_1=X_3\}}$, $Y_3 = 1_{\{X_2=X_3\}}$. Dann sind jeweils Y_1 und Y_2, Y_1 und Y_3, Y_2 und Y_3 unabhängig, aber Y_1, Y_2, Y_3 zusammen *nicht*.

(Es ist $P(Y_i = 1) = \frac{1}{2}$, z.B. $P(Y_1 = 1, Y_2 = 1) = P\big(\{X_1 = X_2 = X_3 = 1\} \cup \{X_1 = X_2 = X_3 = 0\}\big) = (1/2)^3 + (1/2)^3 = 1/4 = P(Y_1 = 1)P(Y_2 = 1)$, $P(Y_1 = 1, Y_2 = 0) = P\big(\{X_1 = X_2 = 1, X_3 = 0\} \cup \{X_1 = X_2 = 0, X_3 = 1\}\big) = (1/2)^3 + (1/2)^3 = 1/4 = P(Y_1 = 1)P(Y_2 = 0)$, etc.)

Man sagt dazu auch: Y_1, Y_2, Y_3 sind *paarweise unabhängig*, aber eben nicht unabhängig.

Bemerkung 1.54 (Fortsetzung)

4. Vergleichen wir Def. 1.49 und Def. 1.51, so sehen wir:

Ereignisse $A_1, A_2, \ldots, A_n$ sind genau dann unabhängig, wenn dies für ihre Indikatorvariable $1_{A_1}, 1_{A_2}, \ldots, 1_{A_N}$ gilt.